International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Classify hypochondriac Situations Using Medication Reviews

Sharda Nand Jha[1], Pranav Kumar Parihar[2]
Dept. of Software Engineering, Delhi Technological university , Delhi, India

---***---

Abstract - Among the most crucial concerns facing society is healthcare for humans. It looks for the most accurate diagnostic and comprehensive illness in order to guarantee that patients get the care they need as soon as feasible. Because this recognition is often difficult, additional fields are needed for the health-related component of searching, such as computer science and statistics[3]. To adopt novel approaches, these disciplines need to go beyond the conventional ones. Since there are so many novel approaches, it is feasible to provide an overview without going into specifics. To this end, we suggest a comprehensive analysis of illnesses associated with machine learning in humans.
This research focuses on current methods related to machine learning growth applied to the diagnosis of human illnesses in the medical business in order to identify intriguing trends, generate insignificant forecasts, and support decision-making[1]. This study explores unique machine learning techniques used in healthcare applications to generate appropriate decision support[4], [5]. This work aims to close the research gap by creating a workable decision support system for medical applications.

***Key Words*: Human disease, Machine learning, Neural Networks.**

1. INTRODUCTION

The recent global shifts have underscored the critical importance of healthcare, prompting innovative solutions to bridge gaps in access and accuracy. Amidst the challenges, virtual doctors have emerged as essential tools, particularly for remote settlements lacking healthcare infrastructure. However, while these virtual platforms offer convenience and accessibility, ensuring their accuracy remains a paramount concern, especially in emergencies where physical examination is not feasible.
Virtual doctors are board-certified healthcare providers that prefer to make appointments over the phone and via video; in an emergency, however, this is not an option[6].

In response, a concerted effort has been made to develop and refine machine learning models for disease prediction, leveraging diverse methodologies like gaussian naive Bayes and decision trees. These models, drawing insights from a plethora of data including drug reviews, age, and gender, hold promise in revolutionizing personalized healthcare by providing accurate prognoses and treatment plans. By harnessing the wealth of information embedded in medication reviews, researchers aim to enhance the predictive efficacy of these algorithms, paving the way for more precise healthcare decision-making

Following that, the data set was processed through a range of machine learning (ML) models, such as gaussian naive Bayes, passive-aggressive classifiers, fine, medium, and coarse decision trees[4], [7]. Each model produced the disease with a varied degree of accuracy since all models processed the data using the same set of input parameters. The model that has the highest level of accuracy has been selected.

This interdisciplinary approach not only taps into the vast potential of machine learning in medical research but also underscores the transformative power of data-driven insights. As the study delves deeper into the predictive capacity of medication reviews and evaluates various algorithms, it seeks to not only narrow existing knowledge gaps but also chart a path towards more effective disease condition forecasting. Ultimately, this endeavor holds the potential to redefine healthcare delivery, empowering both patients and practitioners with timely and accurate insights.

2. LITERATURE SURVEY

Numerous studies have been conducted on the harvesting of adverse medication reactions utilising reviews in social networks, as evidenced by multiple papers on this topic. These studies use a range of methods to identify unfavourable client feedback. The most widely used tactic is the definition-based approach.[1] and [2] Adverse drug responses are compiled from medication labels, clinical test results, and user reviews from health-related communities in dictionaries.

The bulk of papers outline machine learning approaches in their methodology. The authors used the random forest technique and Support Vector Machine (SVM) in Convolutional Fields [6], [7]. The most notable aspects of machine learning include N-grams, voice tag components, semantic categories from larger tasks, the quantity of negated contexts, the belonging lexicon-based characteristics for ADRs, drug names, and word embeddings. Previous studies show that ADR detection results can be improved by applying features based on polarity classification, topic modeling[8], sentiment analysis, and subjectivity analysis.

Recently, encoder-decoder networks for ICD coding, recurrent neural networks for extracting ADRs, and convolutional neural networks for ADR classification have all been employed in various studies[10], [8]. Convolutional neural networks have also been studied by authors to predict sociodemographic factors based exclusively on reviews[5]. There are two main types of works on sentiment analysis for drug reviews: those that apply lexicons with sentiment scores or those that employ supervised classification to learn sentiments.

Among the earliest studies on sentiment analysis of medication reviews Xia et al. developed a topic classifier using patient data using multiple polarity classifiers[3], [4], and [5]. Na et al. present a clause-level sentiment analysis method that considers many review aspects, including overall satisfaction, effectiveness, side effects, and condition. Here, the sentiment polarity of individual sentences is calculated using a rule-based method that takes lexical annotation and grammatical relations into account[7], [8]. Aspect-based sentiment analysis is used to categorise patient comments regarding cancer drugs. Opinion words are identified here by use of a lexical resource, from which general sentiments are produced.

Analyse patient medication satisfaction using sentiment analysis and supervised learning. This study found three levels of polarity when comparing neural network-based and SVM-based methods. Many research have attempted to improve cross-domain sentimental analysis or domain adaptation [1], [2], [9], but not at the level of medicine review elements, but rather among various things such as products, movies, or restaurants. A comprehensive overview of the literature on cross-domain sentiment analysis is provided.

Additionally, research has concentrated on the application of machine learning methods to the prediction of particular human illnesses from medication reviews. Based on patient experiences provided in medication evaluations, [16] created a decision support system that uses decision trees to forecast anxiety and depressive illnesses. The study illustrated how machine learning can help medical personnel diagnose mental health conditions.

The body of research on the subject of using machine learning algorithms to drug reviews-based human disease prediction is quite insightful[3][4]. The research stress how crucial it is to take into account a variety of aspects when creating predictive models, including sentiment analysis, contextual information, and particular diseases.

	Unnamed: 0	drugName	condition	review	rating	date	usefulCount
0	206461	Valsartan	Left Ventricular Dysfunction	"It has no side effect, I take it in combination of Bystolic 5 Mg and Fish Oil"	9.0	May 20, 2012	27
1	95260	Guanfacine	ADHD	"My son is halfway through his fourth week of Intuniv. We became concerned when he began this last week, when he started taking the highest dose he will be on. For two days, he could hardly get out of bed, was very cranky, and slept for nearly 8 hours on a drive home from school vacation (very unusual for him.) I called his doctor on Monday morning and she said to stick it out a few days. See how he did at school, and with getting up in the morning. The last two days have been problem free. He is MUCH more agreeable than ever. He is less emotional (a good thing), less cranky. He is remembering all the things he should. Overall his behavior is better. \r\nWe have tried many different medications and so far this is the most effective."	8.0	April 27, 2010	192
2	92703	Lybrel	Birth Control	"I used to take another oral contraceptive, which had 21 pill cycle, and was very happy- very light periods, max 5 days, no other side effects. But it contained hormone gestodene, which is not available in US, so I switched to Lybrel, because the ingredients are similar. When my other pills ended, I started Lybrel immediately, on my first day of period, as the instructions said. And the period lasted for two weeks. When taking the second pack- same two weeks. And now, with third pack things got even worse- my third period lasted for two weeks and now it's the end of the third week- I still have daily brown discharge.\r\nThe positive side is that I didn't have any other side effects. The idea of being period free was so tempting... Alas."	5.0	December 14, 2009	17
3	138000	Ortho Evra	Birth Control	"This is my first time using any form of birth control. I'm glad I went with the patch, I have been on it for 8 months. At first It decreased my libido but that subsided. The only downside is that it made my periods longer (5-6 days to be exact) I used to only have periods for 3-4 days max also made my cramps intense for the first two days of my period, I never had cramps before using birth control. Other than that in happy with the patch"	8.0	November 3, 2015	10
4	35696	Buprenorphine / naloxone	Opiate Dependence	"Suboxone has completely turned my life around. I feel healthier, I'm excelling at my job and I always have money in my pocket and my savings account. I had none of those before Suboxone and spent years abusing oxycontin. My paycheck was already spent by the time I got it and I started resorting to scheming and stealing to fund my addiction. All that is history. If you're ready to stop, there's a good chance that suboxone will put you on the path of great life again. I have found the side-effects to be minimal compared to oxycontin. I'm actually sleeping better. Slight constipation is about it for me. It truly is amazing. The cost pales in comparison to what I spent on oxycontin."	9.0	November 27, 2016	37

Figure 1: drugs reviews

3. METHODLOGY

From an open-source dataset, we created an excel sheet containing all of the symptoms connected to each illness. Then, based on the disorders, age and gender were determined as dataset components. There were more than 230 conditions listed, each with more than a thousand different symptoms and a thousand medication reviews. Data about an individual's symptoms, age, and gender were input to a variety of machine learning algorithms.

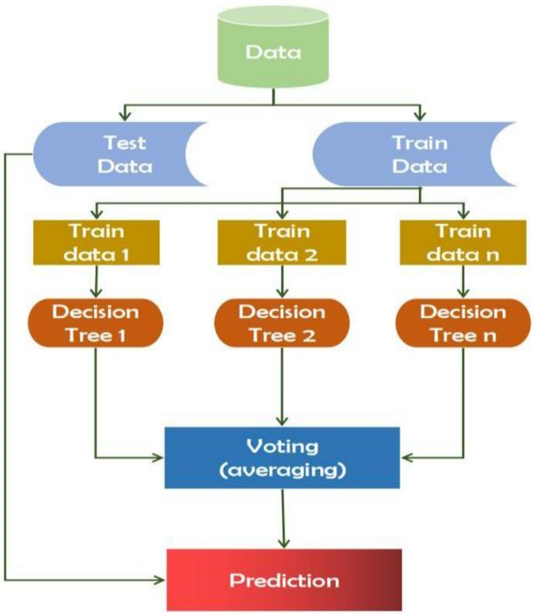

Figure 2: Working mechanism of random forest technique.

3.1 NAÏVE BAYES

It is a machine learning algorithm that uses Bayes' probability theorem to solve classification issues. This is mostly used for text categorization, which requires large training data sets[8].

$$P(B_j \mid A) = \frac{P(A \mid B_j)P(B_j)}{\sum_{i=1}^{n} P(A \mid B_i)P(B_i)}$$

Figure 3: naïve bayes theorm

where P(h|d) is the probability that, in light of evidence d, hypothesis h will hold. This is what the posterior probability alludes to. The probability that data d would exist if hypothesis h were true is denoted by P(d|h). The likelihood that hypothesis h is true (independent of the data) is given by P(h). What we call is the probability prior to h. P(d) is the likelihood of the data (independent of the hypothesis).

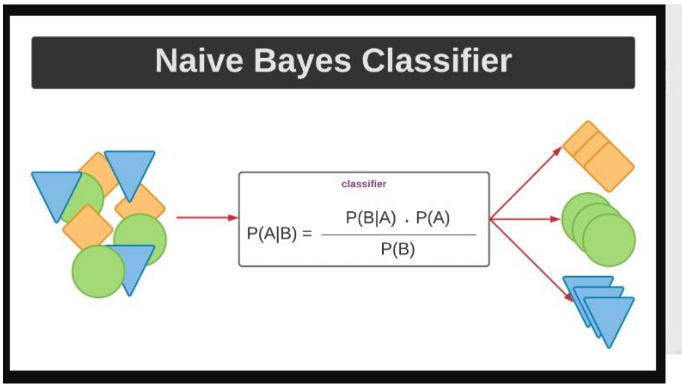

Figure 4 : graph representation of naïve bayes

3.2 DECISION TREES

The decision tree algorithm is a member of the family of supervised learning algorithms. It is used for classification and regression. For prediction, the decision tree's top uses the tree diagram approach. It splits in the prevailing input feature after starting as a root node and splitting again. These steps are repeated until all inputs have been deposited [9], [11], at which point the input is categorised using the weights stored at the very last node. A coarse tree can have up to four splits from each node. On the other hand, in a medium tree, each node can split up to 20 times. A fine tree can have up to 100 splits from each node.

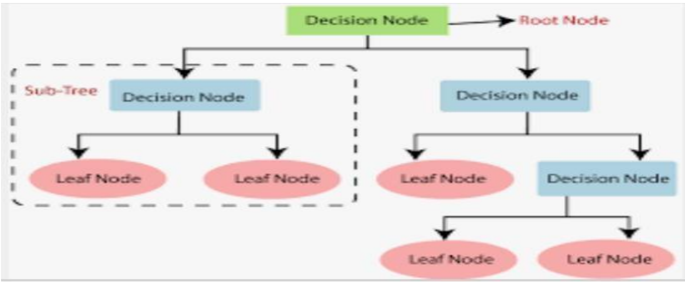

Figure 5: representation of Descision trees

3.3 PASSIVE AGGRESSIVE CLASSIFIER

Passive-aggressive algorithms are commonly used in large-scale learning. This is one of the few "online-learning algorithms". Online machine learning approaches use sequential input data and update the machine learning model one step at a time, in contrast to batch learning, which uses the entire training dataset all at once. This is particularly useful in situations where there is a lot of data and training the entire dataset would be computationally impractical because of its size[11], [12]. In other words, an online learning system will obtain a training sample, modify the classifier, and then reject the sample.

.

If the forecast comes true, passively keep the model in place without making any modifications. Put otherwise, there is not enough data in the case to change the model in any way. Aggressive: If the prediction proves to be incorrect, make changes to the model. Said another way, it might be corrected by changing the model.

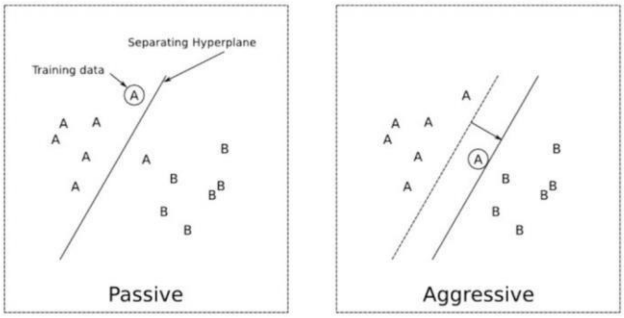

Figure 5 : passive Aggressive

4. RELATED WORK

My initial thought was to use reviews to classify the ailments of patients. Recognising human situations could be important for online systems. Anticipating the reviewer's disposition is the second stage [3]. Sorting drugs for the same condition (in my case, "Birth Control") is the next stage.

Two.csv files containing the data were created; one file was used for testing, and the other for training[7]. Concatenating the two datasets was all that was required to accomplish this, which made it easier to clean and modify the data all at once.
an example of a template formatEven after they have been defined in the abstract, define acronyms and abbreviations whenever they are used for the first time in the text. It is not necessary to define abbreviations like IEEE, SI, MKS, CGS, sc, dc, and rms. If they are not avoidable, do not use abbreviations in the title or headers.

There will unavoidably be some artefacts in text data. At first, I believe I'm looking at apostrophe code in HTML. For data science, I employed the OSEMN methodology, which calls for cleaning (scrubbing) data as soon as it is obtained (presuming there is still more to clean) [9]. I hope there is no misunderstanding because I used three distinct terms (cleaning, scrubbing, and preparation) to describe the same thing.

Preprocessing

The first stage in natural language processing is to adequately clean and transform text. If you don't, your model will function very poorly or not at all.

Taking Out Stop Words

Words such as "I," "was," "there," "me," and so on don't convey much meaning to a neural network; instead, they generate noise and hinder learning.

Lemmatization

To minimise dimensionality and increase classification accuracy, verbs could be changed into their root form since words like learning, learned, and learn all imply the same thing.

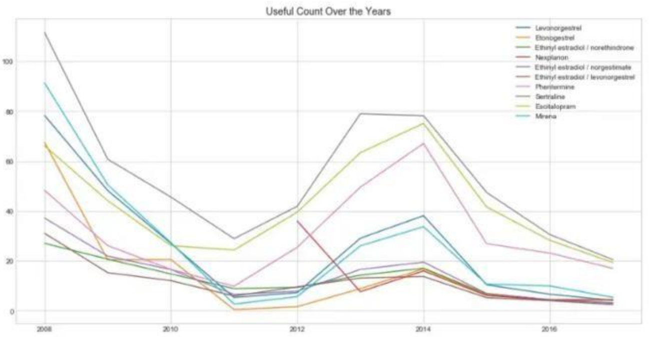

Fig 6 : top most reviewed drugs have similar pattern

5. RESULT

We employ the Passive Aggressive strategy since it is a more advanced method than Multinomial Naive Bayes and counting the test. We depict the confusion matrix for improved visualisation and use accuracy as the evaluation metric.

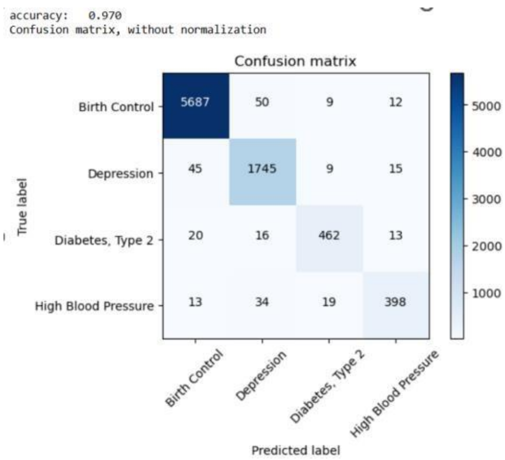

Fig 7 : Confusion Matrix for Naïve Bayes

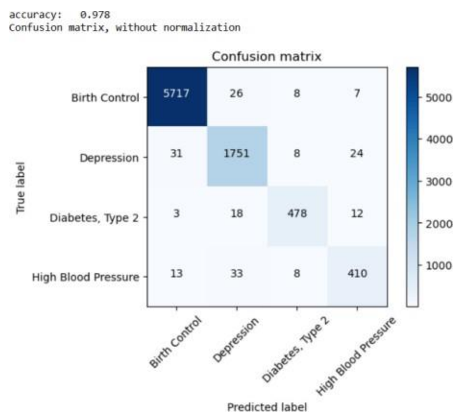

Fig 8 : Confusion Matrix for Passive Aggressive

We obtained 97.8% accuracy by using the Bag of Words model with the Passive Aggressive method. Since birth control classes make up the majority of the classes, the most forecasts are seen in this area.

The outcomes employing evaluation metrics on a bag of words vectorization technique are displayed in Table I. It is evident that the perceptron algorithm performs better than any other categorization techniques. The accuracy of Multinomial Naïve Bayes is 97%, whereas Passive Aggressive achieved an AUC value of 97.8%.

Model	Accuracy
Multinomial Naïve Bayes	0.970
Passive Aggressive	0.978

6. CONCLUSIONS

Using machine learning methods including Naive Bayes, Random Forest Classifier, K-Nearest Neighbours, and Support Vector Machines, this research developed a method of identifying and predicting the presence of a disease in an individual given the user's specific drug review information. Consequently, the proposed methodologies also lead to a comparative analysis of different machine learning algorithms for multiclass classification. Though it is widely believed that the suggested approach may diagnose diseases and identify them while also minimising the cost of medical diagnosis, therapy, and doctor consultation, the accuracy of the disease prediction does depend on the validity of the review.

The passive-aggressive classification achieved the highest disease prediction accuracy, 97.8%, using the previously indicated factors. Almost all machine learning models yielded high accuracy ratings.

REFERENCES

[1] Diana Cavalcanti and Ricardo Prudêncio. 2017. Aspect-Based Opinion Mining in Drug Reviews. In Progress in Artificial Intelligence, Eugénio Oliveira, João Gama, Zita Vale, and Henrique Lopes Cardoso (Eds.). Springer International Publishing, Cham, 815–827M.

[2] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. 2011. Domain Adaptation for Large-scale Sentiment Classification: A Deep Learning Approach. In Proceedings of the 28th International Conference on International Conference on Machine Learning (ICML'11). Omnipress, USA, 513–520

[3] [12] A. Mishra, A. Malviya, and S. Aggarwal. 2015. Towards Automatic Pharmacovigilance: Analysing Patient Reviews and Sentiment on Oncological Drugs. In 2015 IEEE International Conference on Data Mining Workshop (ICDMW)1402–1409. https://doi.org/10.1109/ICDMW.2015.230.

[4] Vinodhini Gopalakrishnan and Chandrasekaran Ramaswamy. 2017. Patient opinion mining to analyze drugs

satisfaction using supervised learning. Journal of Applied Research and Technology 15, 4 (2017), 311 – 319. https://doi.org/10. 1016/j.jart.2017.02.005

[5] K. Kourou, T.P. Exarchos, K.P. Exarchos, M.V. Karamouzis, D.I. Fotiadis, Machine learning applications in cancer prognosis and prediction, Computational and structural biotechnology journal 13, 8 (2015)

[6] A.U. Haq, J.P. Li, M.H. Memon, S. Nazir, R. Sun, A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms, Mobile Information Systems 2018 (2018)

[7] M. Maniruzzaman, M.J. Rahman, B. Ahammed, M.M. Abedin, Classification and prediction of diabetes disease using machine learning paradigm, Health Information Science and Systems 8(1), 7 (2020)

[8] D.R. Langbehn, R.R. Brinkman, D. Falush, J.S. Paulsen, M. Hayden, an International Huntington’s Disease Collaborative Group, A new model for prediction of the age of onset and penetrance for huntington’s disease based on cag length, Clinical genetics 65(4), 267 (2004)

[9] T.V. Sriram, M.V. Rao, G.S. Narayana, D. Kaladhar, T.P.R. Vital, Intelligent parkinson disease prediction using machine learning algorithms, International Journal of Engineering and Innovative Technology (IJEIT) 3(3), 1568 (2013)

[10] S. Vijayarani, S. Dhayanand, Liver disease prediction using svm and na¨ıve bayes algorithms, International Journal of Science, Engineering and Technology Research (IJSETR) 4(4), 816 (2015)

[11] Koby Crammer, Ofer Dekel, Joseph Keshet, Shai Shalev-Shwartz ,Yoram Singer, Online Passive-Aggressive Algorithms ,School of Computer Science and Engineering The Hebrew University Jerusalem, 91904, Israel

[12] S. Karthika , N. Sairam , A Naïve Bayesian Classifier for Educational Qualification, School of Computing, SASTRA University